Slide 1 of 14

Title Slide: Postsecondary Education Data Sample Design, Weights, Variance, and Missing Data

Slide 2 of 14

This module describes the weights and variables associated with standard error calculations for postsecondary education data. It also summarizes the postsecondary education studies' sample designs and describes techniques for calculating estimates and their standard errors. Finally, the module describes how missing data are handled in postsecondary education studies.

Slide 3 of 14

The NPSAS two-stage sampling design involves the selection of a nationally representative sample of postsecondary education institutions, followed by the sampling of students within these institutions. NPSAS is explicitly stratified by institution type – that is sampled institutions explicitly represent all major sectors of postsecondary education, including public, private not-for-profit, and private for-profit; and less-than-2-year schools, community colleges, 4-year colleges; and major universities with graduate-level programs.

These institution types are a combination of control and level. By control we mean public and private, not-for-profit and for-profit. By level we mean less-than-2 year schools, community colleges, 4-year colleges, and doctoral or non-doctoral granting universities. As NPSAS explicitly stratifies by control and level – data are available that allow researchers to generalize to these strata. Caution should be used in generalizing beyond these sampling strata.

Slide 4 of 14

As was discussed in Common Module 3, weights should be used in analyses of data from studies like the postsecondary education studies in order to make estimates produced from the sample representative of the target population. In the postsecondary education studies, the sampling weights are designed to account for unequal probabilities of selection and nonresponse.

Slide 5 of 14

Since NPSAS:04, one weight is included for all study members, while earlier NPSAS studies included separate weights for cases that participated in the student interview versus those with data from any source.

The postsecondary longitudinal studies may have one or more weights. Each weight reflects a defined population represented by the cases with a given source of data and a given point in time. Due to the multiple data collections associated with a longitudinal study, B&B and BPS have more than one weight to reflect the subset of study members

participating at each follow-up as well as an overall panel weight for cohort members who participated throughout the study.

Slide 6 of 14

For example, within the B&B you will find four different weights available for use, the base year weight – which includes the students interviewed at graduation; the base year/3rd follow-up weight – which includes students who responded to the interview at the base year and again at their tenth year after graduation; the panel weight – which includes students who responded to all interviews – first at the base year and at their first, third, and tenth years after graduation; and the 1st follow-up/3rd follow-up weight – which includes students who responded to the interview at their first and tenth year after graduation.

For most analyses, the suggested weight is the panel weight. However, a different weight may be appropriate for your specific research question and may allow for additional cases in your analyses.

Slide 7 of 14

In Common Module 4, two standard calculation procedures were discussed: Replication Techniques and Taylor Series Linearization. Replication is a method that calculates appropriate standard errors based on differences between estimates from the full sample and a series of created subsamples, or replicates. If this is the method you will use to calculate standard errors, you need to select replicate weights that are associated with the analytic weight you are using. For example, the replicate weights WTA001 to WTA200 are the replicate weights for the weight WTA000. Most of the postsecondary studies use the Balanced Repeated Replicates (or BRR) method, which should be specified within the statistical software used for analysis. Always consult the technical documentation for the specific method for the postsecondary education study of choice.

Slide 8 of 14

This page provides an example of the weights provided on the BPS:04/09 data file. The first variable listed, WTC000, is the BPS:04/09 transcript weight. The 200 replicate weights associated with that transcript weight are shown here in a truncated list. The full list of 200 replicate weights that should be used when calculating standard errors using replication techniques is provided on the README file on your restricted-use CD. These replicate weights have a sequential three-digit number following their WTC variable name.

Slide 9 of 14

The second method for adjusting standard errors is the Taylor Series Linearization Method. This method uses PSU and strata identifiers to calculate the appropriate standard errors. If this is the method you will use to adjust standard errors, you need to select the PSU and strata identifiers that are associated with the analytic weight you are

using. The README file on the restricted use CD will provide specific information about the types of weights for each study. Although variables for Taylor series linearization are provided, use of replication techniques and BRR weights is preferred.

Slide 10 of 14

This page shows the same list of weights provided on the BPS:04/09 data file that we just looked when discussing standard error estimation using replication techniques. However, now the weights that are relevant to the Taylor series linearization method for calculating standard errors are highlighted. Now the PSU, stratum, and BPS:04/09 transcript weights are highlighted as they are the weights that should be used for Taylor series methods.

Again the README file provided on the restricted-use CD details all of the variables available for analysis on your data file.

Slide 11 of 14

Across the postsecondary education studies, information may be missing for any item and for any case. This may occur for several reasons, such as when a respondent does not know or does not want to provide an answer to a question or if a respondent ends an interview without completing the survey. Missing data due to nonresponse can be a problem, particularly when there are important differences in the characteristics of respondents who do or not have data on a particular item (for example: if older respondents are less likely to report age, the overall age of the sample may be downwardly biased).

Slide 12 of 14

There are many ways to handle nonresponse and missing data. NCES typically employs statistical imputation techniques when deriving variables for the restricted use files, so data users should encounter very little data missing due to nonresponse – this was discussed in depth in the module titled ‘Data Collected through the Postsecondary Education Studies’ and can be reviewed by clicking on the underlined screen text ‘imputation’ on this page. However, a case may be missing data for reasons other than nonresponse, including for legitimate skips in the questionnaire (for example: a respondent who reports being single will not be asked the question about spousal age). NCES uses reserve codes to label why a case is missing for a variable. Analysts need to pay attention to these reserve codes so they can account for missing data in analyses.

In the postsecondary education data files, -3 is the reserve code that is used for legitimate skips, -9 typically indicates missing, and a reserve code value of -6 denotes a value that was out of range.

Slide 13 of 14

This module has described the weights and variables associated with standard error calculations for postsecondary education data. It also summarized the postsecondary education studies' sample designs and describes techniques for calculating estimates and their standard errors. Finally, the module described how missing data are handled in postsecondary education studies.

Slide 14 of 14

In addition it has provided resources that can be accessed through the DLDT system and/or on the NCES website.

You may now proceed to the next module in the series, or click the exit button to return to the landing page.